INF5442: Image Sensor Circuits and Systems

Soman Cheng, Johannes Sølhusvik

UiO

Abstract.

Table of Contents

Introduction

1 Exercise 1

1.1 Briefly describe the task of each element in a CMOS image sensors signal chain

1. Photons/Scene: The light particles that are captured by the camera are reflected by the objects in the scene
2. Imaging lens: The image lens consists of multiple lens which reflects the scenery to the image sensor. The lenses are used to adjust the focus, removes flare and prevents surface reflection.
3. Microlens array: Microlens are micro sized lenses used to focus the light onto the photodiode to increase quantum efficiency (Q.E)
4. Colour filter array – The colour filter are placed on top of the pixel array to select colour band for each pixel. The most common RGB colour filter used is the Bayer's filter.
5. Q.E: Quantum Efficiency is the term used to describe the percentage of photon detection of a sensor. The ratio used is the number of carriers collected/electrons accumulated by the device to the number of photons of a given energy being detected.
6. C.G: Conversion Gain described the charge to voltage conversion (either in diode itself in case of 3T pixel, or on floating diffusion in case of 4T pixel)
7. SF: Source follower is the built-in amplifier in a pixel. It buffers the input voltage and drive the output line capacitance
8. PGA: Programmable gain stage before ADC
9. ADC: Analogue to digital converter. It converts the analogue input to digital output.
10. BLC: Black Level Compensation remove offsets, such as dark current and ADC offset.
11. DPC: Defect pixel correction. This step is usually done within image processing.
12. LENC: lens vignetting correction
13. CIP/Colour processing: Colour interpolation (demosaicing) to obtain the correct or most accurate RGB values for each pixel position
14. CCM: Colour crosstalk correction by using a pre-calculated matrix. It can also be used for AWB if applicable.
15. TM/Image enhancement: tone mapping to map the image to another device with a different resolution, such as 8bit monitor resolution.
16. JPEG: Compression process.

1.2 How is the energy of a photon related to the wavelength, and what determines the wavelength of a photon

The energy of a photon is inversely proportional to the wavelength. The following equation defines the energy of a photon:

$$E_{photon} = h \times \frac{c}{\lambda} \tag{1}$$

where “c” is the speed of light, “h” is Plank’s constant and λ is the wavelength. Therefore, a smaller wavelength provides higher energy and a longer wavelength provides lower energy. The wavelength of a photon is directly related to the colour of the emitted particle

1.3 What is a micro-lens and what is it used for in image sensors

Microlens is a small lens of a size of micro. They’re manufactured as an integral part of fabrication process and aligned above each photodiode. These are used to increase the sensitivity of the image sensors, increasing Q.E, by concentrating the light into the photon sensing areas, photodiode, and direct light away from the areas that doesn’t need it.

1.4 What does the term "conversion gain" (C.G) mean

Conversion gain is the measure of voltage change caused by a single electron at the charge detection node. It is expressed as:

$$C.G = \frac{q}{C_{FD}} \mu V/e^- \tag{2}$$

Where q is the charge and C_{FD} is the charge-to-voltage conversion capacitance (floating diffusion capacitance)

1.5 How does conversion gain influence light sensitivity of an image sensor

A higher C.G equals a larger threshold/step voltage between two neighbouring/sequential electron values, thus making it more distinguishable about exactly how many photons have hit the sensor. This makes it easier for the ADC of differentiating the levels and the produced absolute voltage difference between the sensor when unsaturated and fully saturated, becomes bigger.

1.6 Suppose a green LED illuminates a 10x10um pixel with $0.5uW/cm^2$ and that the requirement of the sensor’s responsitivity is 50V/sec. If we assume a Q.E of 40%, what will CG have to be in order to achieve the responsitivity requirement

The area of the pixel in cm^2 is

$$\begin{aligned} Area &= (10\mu m \times 0.0001 cm/\mu m)^2 \\ &= 1 \times 10^{-6} cm^2 \end{aligned} \tag{3}$$

The incoming light power at one pixel is

$$\begin{aligned} P_{pixel} &= 0.5 \times 10^{-6} W/cm^2 \times 1 \times 10^{-6} cm^2 \\ &= 0.5 \times 10^{-12} W \end{aligned} \tag{4}$$

Green light has the wavelength of 550nm and by using equation **??**, mentioned in previous task, the energy is then 3.61×10^{-19}J. The number of incoming photons per second is

$$P_{photons} = \frac{0.5 \times 10^{-12}}{3.61 \times 10^{-19}} = 1.38 \times 10^6 photon/s \tag{5}$$

Since the Q.E is 40% and the amount of electron being produces per second is then

$$E = 0.4 \times 1.38 \times 10^6 = 5.54 \times 10^5 electrons/s \tag{6}$$

Which gives a conversion-gain equal to

$$C.G = \frac{50}{5.54 \times 10^5} V/e- = 90.253 \mu V/e- \tag{7}$$

1.7 How many photons per 20msec will hit a $10x10um^2$ pixel that is being illuminated with $1uW/cm^2$ green light (550nm) from a light-emitting diode

The number of photons per second is twice the number calculated in the previous question, since the area is twice as large, it'll be 2.78^6photons/s. For a 20ms period, the number of photons will be

$$Photon = 2.78 \times 10^6 \times 20^3 = 55.6 \times 10^3 photons \tag{8}$$

1.8 If one doubles the lens f-number, what happens to the light intensity on the sensor

Light intensity is inversely proportional to the square of lens F number. If the F-number is doubled, the light intensity reduces to 1/4th of its original value.

1.9 An image sensor at $5x5mm^2$ has an opening angle of 45deg in the diagonal. What is the focal length

Assuming the lens is larger than the sensor itself. The diagonal distance from the centre of the sensor to the corner of the sensor, so half of the diagonal, is then

$$\begin{aligned} y' &= 2.5mm \times \sqrt{2}) \\ &= 3.54mm. \end{aligned} \tag{9}$$

The equation is a basic Pythagoras theorem. Looking at figure 2.4 in the book [1] and using the equation 2.10, we then have

$$\begin{aligned} f &= \frac{y'}{tan(45)} \\ &= 3.54mm \end{aligned} \tag{10}$$

1.10 How does RGB color space differ from YUV

The RGB colour space has the three primary colours red, green and blue and are added together in different proportion to produce an intended colour. In YUV colour space, Y component determines the brightness while U, cyan, and V, magenta, determines the colour. Therefore Y is called the luminance component and U and V are the chroma component. Since human eye is more sensitive to brightness than colour, U and V components can be compressed much greater than Y. This can provide a higher image compression rate without degrading the quality of the image. Removing U and V components renders a grey scale image.

1.11 Convert $[R, G, B] = [200, 187, 50] into [Y, U, V]$ space assuming 8-bit resolution

Y = 0.299*200 + 0.587*187 + 0.114*50 = 175 U = 0.492 * (50 – 175) = -62 V = 0.877 * (200 – 175) = 22

1.12 A blackbody at room temperature (300K) radiates max energy at which wavelength

According to Wien's Displacement Law [2] the maximum wavelength is given by:

$$\lambda_{MAX} = \frac{b}{T} \tag{11}$$

where b is Wien's Displacement constant and T is temperature.

$$\lambda_{MAX} = \frac{2.8977721 \times 10^3 m.K}{300K} = 9.66\mu m \tag{12}$$

1.13 What is the photon flux equivalent to a monochromatic green (550nm) light of 1lux

The Appendix A: Number of incident photons per LUX with a standard light source describes the parameters and equations used to measure photon flux $[photons/m^2/sec]$. By using the equation A.1

$$X_v = K_m \int_{\lambda_1}^{\lambda_2} X_{e,\lambda} \times V(\lambda) \delta\lambda \tag{13}$$

X_v is the photometric quantity which is equivalent of radiance/illuminance and is measured in LUX. K_m is the luminous efficacy for photopic vision and equals to 683lumens/watt. $X_{e,\lambda}$ is the radiometric quantity. V_λ is the photopic eye response and on table A.1 this equals 0.99495 for 550nm, which is green light. By using the equation 13 the radiometric quantity equals

$$\begin{aligned} X_{e,\lambda} = & \frac{1}{683 lumens/watt \times 0.99495} \\ & = 1.47 \times 10^{-3} \end{aligned} \tag{14}$$

From previous task, 550nm equals a photon energy of 3.61×10^{-19}J. Therefore, the photon flux is

$$\begin{aligned} PhotonFlux = & \frac{1.47 \times 10^{-3}}{3.61 \times 10^{-19}} \\ & = 4.08 \times 10^{15} photons/s/m^2 \end{aligned} \tag{15}$$

Source [1]

2 Exercise 2

2.1 Define cut-off wavelength, and what is the value for silicon

The photons hitting the image sensor's pixels must have a certain amount of energy to free electrons from their valence band into the conduction band region. Given that the energy of a photon is dependent of the wavelength of the photons and the maximum wavelength that is capable of exciting the electron is called the cut-off wavelength, λ_{cutoff}. Generally it means the photodiode can sense up to this wavelength, any wavelength larger than $\lambda_{cuttoff}$ will not generate electrons in the pixel and instead pass right through. In silicon $\lambda_{cuttoff}$ is approximately $1.1\mu m$.

2.2 50% of red, green, and blue light in silicon is absorbed at which depth

The light penetration depth in a certain material is defined by the absorption coefficient and is dependent of the wavelength. Note that red light is defined with 600nm wavelength, green with 550nm and blue with 450nm. The sensing material of interest is silicon and so the graph 1 can be used to map the where 50% of these wavelength are absorbed.

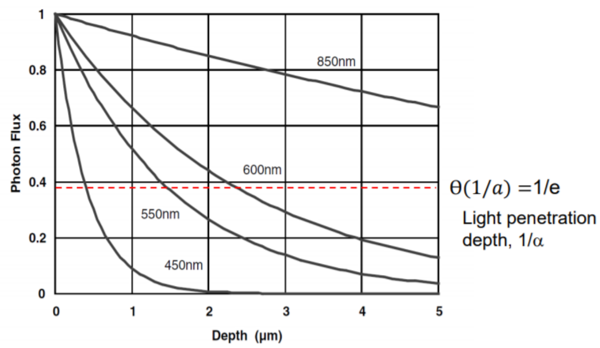

Fig. 1. Absorption of light in silicon. Source [1]

A rough measurement gives $depth_{red} = 1.75\mu m$, $depth_{green} = 1.15\mu m$ and $depth_{blue} = 0.33\mu m$.

2.3 Is it possible to only read out a small portion say 10x10pixels window (region-of-interest) of a larger CCD sensor? How is this different from a CMOS sensor

It's not possible to read a specific section of a CCD image sensor, as all the pixels are read simultaneously and shifted out at once due to it's shift registers.

CMOS image sensors, on the other hand, can read a small section of pixels. They use row and column decoders which makes it possible to select a specific row and then the column of interest, so it's still restricted and can't read the pixels outside of the range of rows or column being selected.

2.4 Why are CCD sensors limited to analogue output, only? Why not integrate A/D converters and digital signal processing circuits

CCD fabrication process differs from the general CMOS based process and is therefore inefficient in terms of cost and performance. The transistors produces for digital circuit usage are poor and is more beneficial and efficient to separate these. As CCD image sensors are large in size due to their shift registers, including digital circuit parts will only increase the total die size of the chip.

2.5 Explain the difference between global shutter and rolling shutter readout

In a global shutter the start and end of the integration time is the same and happens simultaneously for the entire pixel array. This means the whole image is captured at the exact same time with no delay. In rolling shutter the integration time is the same for all pixels, but starts and ends at a different time, often with a small delay between each row and hence the given name rolling shutter. The integration time begins with a single row and and the next one begins when it's done and the read out process begins. This means the scene is being captured row by row.

2.6 What are the pros and cons of CCDs versus CMOS image sensors

CCD		CMOS	
Pros	**Cons**	**Pros**	**Cons**
Has global shutter and prevents distortion	Low resolution and large die size	Low power consumption	If running on rolling shutter, it suffers from distortion and artefacts
High signal to noise ratio	The manufacturing process is expensive	Has integrated digital circuits	Less sensitive to light
	High power consumption due to shift registers	Cheaper to produce	

2.7 Explain the artefacts that can occur when rolling shutter sensors capture fast moving objects

As a rolling shutter captures the scene rows by row, there is a certain delay between each row and this delay can cause

- Skew: Diagonal bend as the camera or object moves, capturing parts of the object at different times.
- Smear: This artefact appears when something is rotating quickly (propeller). The smear of each blade is caused by the propeller rotating at the same or near the same speed that the frame is read by the camera.

Other artefacts a rolling shutter suffers from are:

- Partial Exposure: If a flash goes on only partial of the exposure time, the flash may only be present at certain rows of the pixelarray in a given frame.
- Wobble: This phenomenon occurs when the camera isn't stable, but vibrating. The resulting image will appears to wobble and is blurry.

2.8 Why do most CMOS image sensors use rolling shutter instead of global shutter

A rolling shutter sensor will have less noise and a wider dynamic range than a global shutter. To apply global shutter, the image sensor will need an additional space for saving the actual signal accumulated by light, such as an internal capacitor separated from the photodiode or external memory. Both schemes requires additional space and increases the chip size, also an internal capacitor will reduce the fill factor.

2.9 Let a 10x10um ideal photon detector be illuminated by 10k photons. What is its signal/noise ratio

Since no read noise value is presented in the task, it is assumed photon shot noise is dominant. Therefore, the equation 3.47[1] can be used

$$\begin{aligned} SNR &= 20log\sqrt{N_{sig}} \\ &= 20log\sqrt{10k} \\ &= 40dB \end{aligned} \tag{16}$$

3 Exercise 3

3.1 Define conversion gain

Conversion gain is the measure of voltage change caused by a single electron at the charge detection node. It is expressed as:

$$C.G = \frac{q}{C_{FD}} \mu V/e^{-} \tag{17}$$

Where q is the charge and C_{FD} is the charge-to-voltage conversion capacitance (floating diffusion capacitance)

3.2 A 2.2um 4T pixel has maximum output voltage swing of 1.1V, and FWC is 14ke-. What is the conversion gain? You can assume source follower gain of 0.8

The maximum voltage output swing is the value after the amplification

$$V_{C.G} = \frac{1.1V}{0.8} = 1.375V \tag{18}$$

To find the conversion gain, the equation for voltage swing at the FD can be used

$$\varDelta V_{FD} = C.G \times S_{electrons} \tag{19}$$

by rearranging the formula, the C.G can be found

$$\begin{aligned} C.G &= \frac{\varDelta V_{FD}}{S_{electron}} \\ &= \frac{1.375V}{14000e-} \\ &= 98\mu V/e- \end{aligned} \tag{20}$$

3.3 If temporal noise floor is 2.3e- rms, what is the dynamic range of this pixel

No parameters are given, therefore parameters from the previous task are used.

$$\begin{aligned} DR = 20log\frac{FWC}{n_{noise}} \\ = \frac{14000e-}{2.3e-} \\ = 75dB \end{aligned} \tag{21}$$

3.4 Why does Q.E for short wavelengths eventually drop to zero

The photons entering the pixel and hitting the photodiode, which generates electron-hole pairs by the photon energy. The depth at which these pairs are generated depends on the wavelength of the photon. When the wavelength is too short, the photons will only manage to reach the surface of the pixel. In this case the energy is either absorbed in this area or reflected and as a result, photons of short wavelength do not reach the depletion area. Therefore, no charge are generated due to short wavelength and the quantum efficiency drops to zero.

See figure 3.5

3.5 Why does Q.E for long wavelengths eventually drop to zero

If the photon energy is not sufficient to generate electron-hole pair, the photon will only pass through the semiconductor. Photon energy is inversely proportion to its wavelength. At higher wavelength, the photon energy is not sufficient enough to excite the electrons from it's band. Hence, the quantum efficiency drops to zero.

See figure 3.5

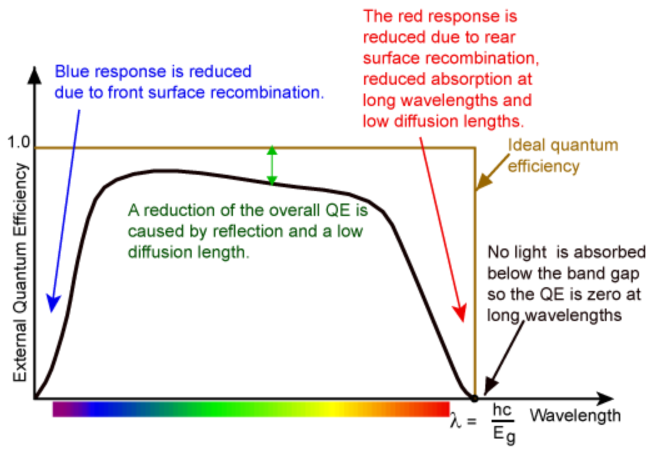

Fig. 2. Q.E graph

3.6 Why does Bayer RGB pattern have twice as many green pixels as red and blue

Human eye perception is more sensitive to the colour green compared to red and blue. By taking advantage of this property, the Bayer RGB pattern adds more green pixels to the colour filter to mimic the human visual system. The resulting image will appear less noisy and has finer details compared to a filter with equal quantities of RGB.

3.7 Can you list three types of fixed pattern noise sources

- Dark current variation – This is due to the charge integrated when pixel is not exposed to light. The reasons for this are the leakage current in the transistors and the electrons generated due to temperature.
- Photo-response non-uniformity – The voltage amplification at pixel amplifiers is not uniform because of process variation. Hence, the signal generation at pixels are non-uniform.
- Vertical FPN – Noise created due to process difference in each transistor and creating an offset for each column during readout.

3.8 List three types of temporal noise sources

- Thermal noise – This noise occurs due to the movement of electrons within resistance due to the temperature and is always present above 0 K.
- Photon shot noise – This is due to the inherent natural variation of the incident photon.
- Flicker (1/f) noise – This noise is due to the surface states that occur due to abrupt discontinuity in semiconductor lattice. These states are caused by dangling bonds at the surface. They combine with the charge and contribute only in low frequency.
- Reset (kTC) noise – This is caused by the MOS switch used to reset the floating diffusion capacitance. This comes from the thermal noise of the MOS switch resistance and is sampled and held by the capacitor, which adds to the signal.

3.9 Explain how FPN can be removed in pictures

FPN can be removed from pictures with Black Level Subtraction technique. A picture is first captured with the shutter closed and thus a dark image is obtained. This dark image will contain the offsets of the pixels at the specific exposure time and temperature, varying these factors can change the FPN. A picture of the desired scene is then captured with the same exposure time and under same temperature and by subtracting the dark image from it will remove all the offsets, producing an image free of FPN.

3.10 Explain how temporal noise can be removed in pictures

Temporal noise varies in time and is dependent of the environment. A simple method to reduce it's effect can be accumulating more signal by increasing the exposure time and hence increasing SNR. Dark current can be reduced by reducing the size of the pixel and using pinned photodiode. Thermal noise and flicker noise can be reduced by optimising the amplifier design by adjusting the (W/L) ratio of the amplifier. Correlated Double Sampling technique helps reduce both kTC noise and flicker noise. This technique samples both the reset value and the signal and taking the difference between the two. This will remove the noise present in the pixel and providing the actual voltage drop regardless of the reset value. Thus the reset value needs to be the same as the one the signal is dropping from.

3.11 Calculate the standard deviation of the number of photo-electrons accumulated in a pixel whose average (mean) value is 1000e-. Assume only photon shot noise. What is the signal/noise ratio of the pixel

Photoelectricity is random by nature and follows the Poisson's probability distribution. Following the Poissonian process, the variance value is equal to the mean value, μ, and the standard deviation, σ, is the the square root of the variance. Therefore, the standard deviation is

$$\begin{aligned} \sigma &= \sqrt{\mu} \\ &= \sqrt{1000} \\ &= 31.623 \end{aligned} \tag{22}$$

As the task mentioned to only assume photon shot noise means the number of electron generated is equal to the average value, 1000e-. Using the formula for SNR

$$SNR = 20log\frac{\mu}{\sqrt{\mu}} = 30dB \tag{23}$$

3.12 If "noise" in a sensor is generally considered to be random (non-deterministic) deviation from its mean value (average value), explain why a "fixed" (deterministic) pattern in image sensors is considered to be "noise"

Fixed pattern noise in image sensor is only fixed in spatial domain and not in time domain. It means that the intensity of the pattern obtained does differs if the environment changes, such as the temperature, exposure time or illumination source. In this case the noise is not deterministic and hence, fixed pattern noise is considered noise. Another reason is that the fixed pattern noise varies from a sensor to another due to process variation and no sensors has the exact same fixed pattern noise despite the images are taking with the exact same condition. Last, but not least, noise deteriorates an image and it's quality and is therefore considered as a noise component.

4 Exercise 4

4.1 What is meant by 'black level' in a digital picture

Black level is the darkest value in an image with no illumination source present. Due to noise, the output of the pixels are rarely zero and produces a non-zero value from the ADC and is known as the black level. This level can be measured by either capturing an image in total darkness or using the optical black pixels.

(I'm not sure if the answer is correct. According to the slide on lecture 4, page 7, ADC output values non-zero even if pixel output signal is zero(??) and is therefore necessary to add an offset in the ADC input to achieve a certain level and subtracts afterwards. As I understood it, this level is the Black Level Compensation.)

4.2 List possible reasons why a digital camera has a non-zero black level

The main cause for a non-zero black level in digital cameras are due to dark current, offset voltage from the PGA and read noise from the ADC. Dark current arises due to thermal agitation of electrons in the analogue circuits and current leakage (see previous exercises).

4.3 Explain why black level must be subtracted before being processed in the signal processing data path inside a camera. Give an example what can happen

Black level is an undesired offset which affects the the resulting image by producing false colour and illumination representation of the scenery. During digital signal processing, such as colour interpolation and auto white balance, the additional offset value will cause a poor approximation of the actual colour in the scene. A non-zero black level limits the usage of the linear region, affecting the dynamic range, which affects the quality of the image and deceasing the dynamic range. Instead of starting from 0, the linear region begins from the offset value and causes the bright values to saturate earlier, hence losing information in the bright region.

4.4 What is the role of the demosaicing (aka colour interpolation) algorithm

Image sensors have a colour filter array above it's sensor, typically in a Bayer's pattern. This causes the pixels to capture only a single colour, either red, blue or green, while the information of the other colours are absent. In order to reconstruct the image with the correct or decent representation of the scenery, these pixels needs the other two values to recreate the correct colour. This is done by using colour interpolation algorithms. The process finds the actual colour by measuring the neighbouring values and calculates an approximation or average value by either using "Nearest Neighbour Interpolation", "Bilinear" or "(Bi)Cubis Interpolation".

4.5 What artifact(s) can demosaicing introduce in the image? What, if anything, can be done to mitigate such issue(s)

The artefacts demosaicing introduces occur when the spatial frequency of the scenery is higher than the resolution of the image sensor. The demosaicing process will struggle to find the proper colour representation of that area. There are two main artefacts: Misguidance Colour Artefacts and Interpolation Artefacts[3].

Misguidance colour artefacts are "False colour" and "Zipper effect". False colour effect occurs when the process struggles to find the proper colour and assigns a false colour around the area, typical near the edges or fine details. Zipper effect produces an abrupt change in intensity or colour, mainly near the edges, due to difficulty of estimation.

Interpolation artefacts are related to the limitation of the interpolation algorithm itself and is less noticeable.

Or...

The artefact demosaicing introduces is aliasing. Aliasing occurs due to spatial high frequency which the sensor is incapable of resolving and colour interpolation enhances the effect. This is due to the details found in the scenery are too small causing a single pixel to capture it or the change of colour, creating discolouration or interference in the resulting image. This can be seen as incorrect colour, intensity or pattern which doesn't exist in the scenery itself.

To mitigate demosaicing artefacts, an optical low-pass filter can be included to reduce the effect or increase the resolution of the image sensor. The use of higher order or more complex algorithm can also reduce the effect, but affects the speed.

4.6 Explain the principle role of the colour correction matrix in a digital camera

Colour Correction Matrix (CCM) is created to compensate for colour cross-talk between the pixels. Along with colour filter, each pixel registers only a single spectral, but the filter is not ideal due to process variation. This causes charge accumulation, the photodiode absorbs photons despite not being the desired wavelength. CCM coefficients defines the level of colour cross-talk between pixels and adjust the R, G, B values to compensate.

4.7 If a CCM equals a unity matrix with only 1s in the diagonal and 0s all other coefficients, what does it say about the sensor spectral response

If CCM equals a unity matrix with only 1s in the diagonal and 0s coefficients

4.8 Explain why large CCM coefficients outside the diagonal result in noisy images

4.9 Can the CCM matrix be adjusted to compensate for changes in the scene illumination spectrum? If yes, explain how

5 Exercise 5

5.1 Can you think of reasons why most tone mapping curves use high gain in dark region and low gain in bright region of the picture

Tone mapping are required to map images from higher resolution devices to those with lower resolution, typically 12bits to 8bits. As human is more sensitive to variances within the dark region compared to the bright region, by using high gain at the dark region and low gain in bright region the tone mapping curve takes advantage of the way human perceives light. Another reason to use low gain at the brighter region is to prevent the values from saturating, hence losing the information accumulated by natural illumination. If linear mapping is used, the resulting image will have low contrast and less pleasant to see.

5.2 Assume a video camera is capturing a scene where the sun is about to disappear behind a cloud; hence more brightness is needed in the picture. What should change first, integration time or gain, and explain why made that choice

The integration time should always be the first factor to increase. This allows the image sensor to accumulate more charges and resulting in an image with a better quality and higher SNR. By increasing gain, the noise will be amplified as well, which reduces SNR.

5.3 Make a flowchart diagram of an auto-focus algorithm

Assuming that this task is requesting for a general auto-focus flowchart diagram.

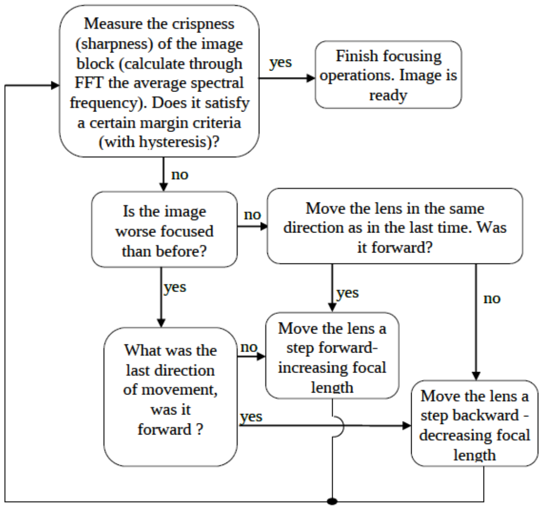

Fig. 3. A simple flow chart diagram of adjusting the lens for auto-focus

5.4 Explain the pros and cons of linear versus cubic interpolation schemes in CMOS sensors

	Linear Interpolation	(Bi)Cubic Interpolation
Pros	- Requires a minimum of 2 points - Simple to implement, faster and requires less power	- Provides a more accurate colour representation - Compared to linear interpolation, cubic interpolation produces a smoother function curve
Cons	- Less accurate for non-linear - Produces a less smooth function curve	- Advance computation and requires more processing power - Requires a minimum of 4 points

6 Exercise 6

6.1 List the three data reduction concepts used in JPEG compression

- Sub-sampling chrome information
- Discrete Cosine Transformation (DCT)
- Quantization
- Run Length Coding(RLC)
- Entropy Encoding

6.2 Why does JPEG use YCbCr instead of RGB data

Human visual system is more sensitive to variance in luminance compared to chromiance. To adapt this property, YCbCr is a preferable scheme. It separates Y, luminance, from the chroma, Cb and Cr, providing the opportunity to work with these individually. Further, the chroma information is then reduced by removing partial of these components without actually reducing the quality of the image.

6.3 Why does JPEG group image data into blocks of 8x8 pixels

Transformation from spatial to frequency domain is performed in JPEG for energy compaction i.e. limited number of transformed coefficients carry most of the signal energy. This requirement is met when the pixels in the average block are correlated in spatial domain. An 8x8 block has a high correlation between pixels for energy compaction in the transformed matrix. It's proven through studies to be the optimal size for computation purpose, requiring less memory space and inexpensive hardware implementation.

A smaller block size can struggle to capture the important pixel-to-pixel correlation. A larger block sizes can be too big, containing uncorrelated pixels and requires higher computation complexity.

6.4 What is the purpose of DCT in JPEG

Dicrete Cosine Transformation (DCT) transforms the micro blocks from the spatial domain to the frequency domain. This is done to find the high frequency components and to further discard these. The resulting block consists of a single DC coefficient, the largest value, at the upper-left corner and 63 AC coefficients for each frequency. This process concentrates the signals in one corner and provides a more effective compression later.

6.5 What is the purpose of quantization in JPEG

Quantization process takes the 8x8 micro blocks, produced by the DCT, and divides them by using a quatization matrix, a lowpass filter. The elements in the matrix controls the compression ratio, where a larger values increases the compression rate and vice versa. The values are then rounded to nearest integer and higher frequency components are rounded to zero.

The main purpose is to achieve smaller positive or negative values, which requires fewer bits to represent, and removing the high frequency components by setting these to 0. Human visual system can't distinguish the exact strength of a high frequency brightness variation and therefore this operation does not affect the resulting image.

6.6 What step(s) makes JPEG compression lossy

Sub-sampling is lossy since partial of the chroma information is discarded. The quantization is considered to be the most lossy operation in the whole process because values are rounded and it's irreversible.

6.7 What is the basic concept used in entropy encoding schemes such as Huffman encoding

Entropy encoding, a lossless data compression, involves arranging block in a zigzag pattern by employing run length coding(RLC) algorithm and further compressing it by allocating bits to the resulting code. The RLC algorithm groups the same frequency and because the data inside the block has high correlation the frequency reoccurs throughout. The concept of entropy encoding is to sort the frequency groups in terms on occurrence and allocate few bits to those with frequent occurrence and longer bits to the rare.

7 Exercise 7

7.1 When a CMOS image sensors outputs pixel data, how does the receiver know which position in the array the pixel value corresponds to? What additional output signals from the sensor are used to help aligning the pixel position?

Assuming the addressing signals are not accessible within the system, the receiver needs to know which pixel the respective data belongs to. There are various method, but the course will use the OV7670 camera module as an example.

The outputs of the CMOS image sensor consist of: VSYNC HREF PCLK

These are the necessary signals to determine which pixel the data belongs to.

Source [4]

7.2 Draw up a timing diagram that illustrates the pixel output timing from a CMOS image sensor with DVP interface

Assuming the interface is the same one as mentioned in previous task

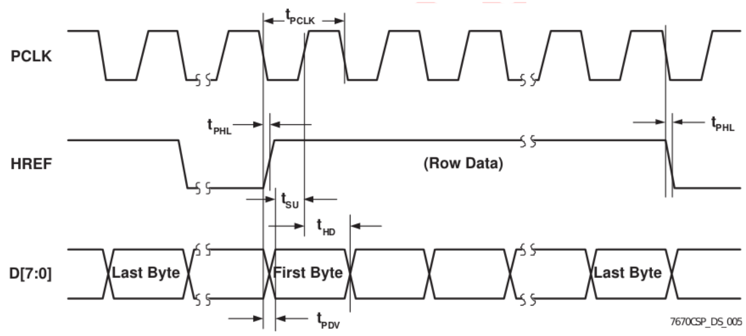

Fig. 4. Timing diagram for HREF

7.3 Explain why two sensors capturing stereo pictures must stay time synchronised and how this is achieved?

Stereo pictures are recorded with two cameras placed at different angles to achieve the 3D effect by capturing the scene from a different point of view. To obtain the best results the cameras need to be synchronised and record the frame at the same time. This is especially true if they're in motion or under unstable illumination. The consequences of not having them synchronised are distorted scenery, objects are shifted and an offset is introduced in the pictures, making it impossible for stereoscopic usage. The software for creating 3D image is dependent of the angle between the two sensors to find the distance from them

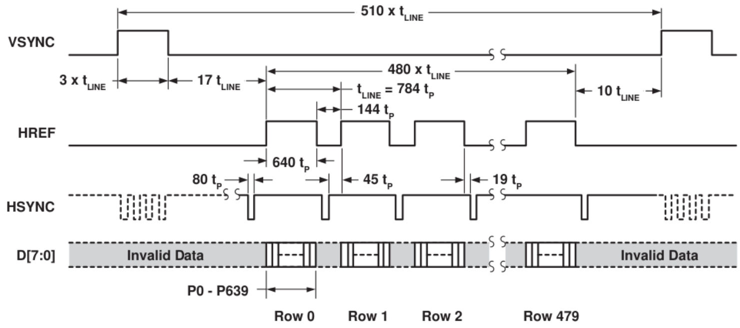

Fig. 5. Readout timing diagram for a "VGA" (640 x 480)

to the actual scene. The distance is then used to pinpoint a specific position in the scene found on both sensors and ensure that the given points are the exact same point in the real scene.

The two cameras can be synchronised by using a common clock source and/or shared trigger for capture. It should be noted that they should share exposure time and settings to ensure maximum cohesion between the resulting images from the sensors.

7.4 What is a pull-up resistor? What purpose does such a circuit serve?

A pull-up resistor is a resistor connected between pins, such as MCU or IC, and VDD. The common impedance for a pull-up resistor is 10K or 100KOhm. This is used to set the floating line to a known state by "pulling" it up to VDD. The MCU or IC can tell whether the line is active or inactive by seeing the state the line is in. Another purpose of this component is that it combats induced noise created by magnetic-fields.

7.5 A sensor outputs uses 12b DVP output with 100MHz pixel clock. Load on output pins is 20pF. VDDIO=1.8V. Calculate worst case current spike during transition and explain how this can result in image noise. Calculate the average current assuming 50% toggling rate. Why is high power a concern? Can you suggest a method to reduce this power?

Considering the "worst case" is when all signals toggles at the same time and from task 1, the output signals for determining the pixel's position needs to be taken into consideration. That would make a total of 15 signals: 12b DVP, clock, HREF and VSYNC.

$$
\begin{aligned}
I_{VDDIO} = \frac{C_{load} \times V_{VDDIO}}{\frac{1}{f_{clk} \times 4}} \times N_{DVPlines} \\
= \frac{20 \times 10^{-12} \times 1.8V}{\frac{1}{100 \times 10^6 \times 4}} \times 15 \\
= 216mA.
\end{aligned}
\tag{24}
$$

The average current, one needs to find the total power consumption and divide it by the voltage. The task mentioned 50% toggling rate, so all signals runs at half of the clock frequency. Although, the clock frequency remains the same.

$$
\begin{aligned}
P_{IO} = C_{load} \times V_{VDDIO}^2 \times f_{clk} \times N_{DVP} \\
= (20 \times 10^{-12} \times 1.8V^2 \times 50 \times 10^6 \times 14) + (20 \times 10^{-12} \times 1.8V^2 \times 100 \times 10^6 \times 1) \\
= 51.84mW
\end{aligned}
\tag{25}
$$

The average current is then

$$
I_{avg} = \frac{0.05184W}{1.8V} = 28.8mA \tag{26}
$$

High power consumption is of concern due to it contribution in increasing the temperature of the chip and can cause damages to partial of the circuits. Large current spikes can induce supply and GND noise and cause variation at GND. The unstable GND will affect the other parts of the circuit. The same occurs for VDD as well and the power supply voltage can drop.

7.6 Why is the analogue supply voltage higher than the digital supply in most CMOS image sensors?

A higher voltage supply will provide analogue circuits a wider linear range to work with and improves it's performance. It increases SNR, output swing and gain. Digital circuits do not require the same amount due to it only needs sufficient voltage level to differentiate 2 states; "0" and "1". A lower voltage supply for digital circuits reduces power consumption and provides faster transitions/switch.

7.7 Why is it important to keep the voltage supplies as low-noise as possible in CMOS image sensors?

In analogue circuits, a stable power supply is necessary to maintain its performance. In image sensor, the supply is connected directly to the photodiodes and the pixels outputs. A noisy power supply will affect these part directly in terms of reset level and incorrect output values. Another important aspect is the ADC. ADC uses the power supply as it's main supply and reference signal. Therefore, if the supply is noisy it'll create an undesired offset which will cause misinterpretation of values from the pixel array.

7.8 Why is external I/O supply voltage (DOVDD or VDDIO) typically separated from the internal digital supply voltage (DVDD)? Can the two values be different? If so, how is this handled inside the chip?

External I/O voltage supply has a higher voltage, 3.3V to 5V, compared to the internal digital supply voltage, which has 1.2V to 1.8V. As mentioned in previous task, digital voltage supplies does not need to be any higher since the value only needs to be high enough to differentiate "0" and what is known as "1". I/O pad requires a higher level due to it needs to drive the output capacitance and ESD circuits. Another advantage of having separate supply is to reduce noise. I/O pads are often exposed to large currents and voltages, affecting it's own supplies and increases the temperature.

To achieve this a DC-DC converter, level shifter and a simple voltage divider can be used.

7.9 Explain how CMOS I/O pins are ESD protected

A simple ESD circuit consists of two reverse biased diodes. Depending on the polarity of the voltage, either too high or low, the upper diode will conduct or the lower diode will conduct respectively and divert the voltage away from the input. In some cases, current limiting resistors are included to prevent the diode form burning out.

7.10 What does "tri-state of I/O pins" mean? Why is this concept used?

A tri-state I/O pins has 3 states: logic high, low and high impedance. When the pin does not receive any input it will be in high impedance state and logic high or low otherwise. This allows the pin to be connected to multiple transmission line and reduces the quantity of I/O pads.

7.11 List at least three reasons why the CMOS image sensor industry is starting to move away from parallel output and over to serial output

7.12 Draw a conceptual diagram of LVDS sender/receiver and list the reasons why this has become such a popular industry standard

- LVDS uses differential signals and is immune to common mode noise. This makes LVDS less sensitive to environmental noise and reduces the risk of noise related problems, such as crosstalk from neighbouring lines. As a result, LVDS can use a lower voltage swing compared with single-ended schemes.

Reduce noise	The amount of ADC reduces to a single one and removes any offset which may occur if using multiple ones
No synchronisation	The receiver will no longer need to align the outputs, but can instead read the signals one after another in serial outputs.
Longer wires if LVDS is used	Due to LVDS properties, the transmission lines can be longer without being affected by noise. Although, one should note that impedance increases with longer wires and should be matched correctly to prevent reflections.
Low power	Less power is required to drive a serial output compared to parallel outputs due to capacitive properties and switching.
Reduce EMC	Toggling multiple lines creates high current spikes and causes high EMC. Therefore, a single output will reduce EMC.

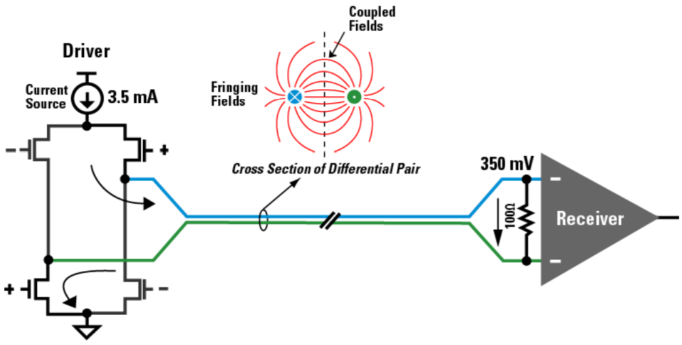

Fig. 6. LVDS

– The differential signalling also results in reduced amount of noise emission. When the two adjacent lines of a differential pair transmit data, current flows in equal and opposite directions, creating equal and opposite electromagnetic fields that cancel one another, hence reducing and magnetic field-induced noise.

– LVDS consumes relatively lower power, because it uses lower voltage swing and draws a constant current. This prevents any current spikes and any noise effect which may occur in the power supplies.

– Lower voltage swing enables LVDS to switch states faster and with a lesser slew rate. This allows a faster bit rate on LVDS i.e. higher operating frequencies.

7.13 A CMOS sensor has 4+1 (4x data+ 1x clk) LVDS output lanes. Calculate the estimated power consumption

Using the parameters found in lecture notes for an LVDS interface lane of 3Gb/s data transfer, each data lane would require 8.75mW and the clock rate needs to be 2 times faster than the data.

$$\begin{aligned} P_{tot} &= (4 * 8.75mW) + (1 * 2 * 8.75mW) \\ &= 52.5mW \end{aligned} \tag{27}$$

Bibliography

[1] Junichi Nakamura. *Image sensors and signal processing for digital still cameras.* CRC press, 2016.

[2] Wien's displacement law. `https://en.wikipedia.org/wiki/Wien%27s_displacement_law`. Accessed: 2017-02-12.

[3] Ruiwen Zhen and Robert L Stevenson. Image demosaicing. In *Color Image and Video Enhancement*, pages 11–22. Springer, 2015.

[4] Jorge Aparicio. Hacking the OV7670 camera module. `http://embeddedprogrammer.blogspot.no/2012/07/hacking-ov7670-camera-module-sccb-cheat.html`, 2012.

[5] D. M. Greenberger, M. A. Horne, and A. Zeilinger. Going beyond Bell's theorem. In M. Kafatos, editor, *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, pages 69–72. Kluwer Academic Publishers, 1989.

[6] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1(3):195–200, 1964.

[7] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777–780, 1935.

[8] N. D. Mermin. *Boojums all the way through.* Cambridge University Press, 1990.